PART I: HIV-1 CTL EPITOPES

SUMMARY

Part I includes tables, maps, and alignments of HIV-specific CTL epitopes arranged sequentially according to the location of the proteins in the HIV-1 genome. We attempted to make this section as comprehensive as possible, requiring that the epitope be contained within a defined region of a maximum of 30 amino acids, but not that the precise boundaries be defined. For more recent updates and useful searching capabilities, please see our Web site: http://hiv-web.lanl.gov/immuno. For concise listing of the best defined CTL epitopes, see the summary by C. Brander and B. Walker in part IV. The same epitope can have multiple entries, as each entry represents a single publication.

TABLES:

Each CTL reference has a six part basic entry:

- **Location:** The amino acid positions of the epitope boundaries and the reference sequence are listed as given in the primary publication. Frequently, these positions as published are imprecise, and do not truly correspond to the numbering of the sequence, but they provide a reasonable guide to the peptide's approximate location in the protein. Also, in many cases the reference sequence identification was not provided, and in such cases it is not possible to use these numbers to specify precise locations.
- **WEAU Location:** The viral strain WEAU (GenBank Accession Number U21135) is used as a reference strain throughout this publication. The position of the defined epitope location on the sequence of the WEAU protein is indicated. Obviously WEAU may not be identical to a given defined reactive sequence, so we are simply indicating the location of the aligned positions. The WEAU numbering is used in the protein maps in this database. Nef in the WEAU cloned sequence has a frame shift, but the Nef reference protein sequence was completed past the frame shift stop codon for the purpose of mapping the epitope locations.

  WEAU was chosen as the reference clone because it is one of the best characterized sequences currently available. The sequence was graciously provided prior to publication by George Shaw. The clone was obtained from a co-culture of this patient's PBMC's, first with normal donor PHA-stimulated lymphocytes for 14 days, and then with the H9 T-cell line for another 14 days. The blood specimen was obtained 15 days after the onset of clinical symptoms of acute (primary) infection, and 35 days after a single sexual encounter (receptive anal intercourse) with a partner whose virus was proven phylogenetically to be responsible for the transmission event. The single nucleotide deletion in nef in the WEAU 1.60 clone is *not* present in the patient's uncultured PBMCs where instead there is a "T." Thus, in the clone WEAU 1.60 nef is disrupted, but in the patient, the virus contains an intact *nef* gene in 10 out of 10 clones analyzed by PCR sequencing. The patient from whom WEAU 1.60 was derived is identified as "Patient #1" in *N Engl J Med* **324**:954–960, 1991 and as "WEAU 0575" in *Science* **259**:1749–1754, 1993. WEAU 1.60 and the virus isolate from which it was derived are SI (syncytium-inducing) strains. The full-length WEAU 1.60 provirus has been sequenced in its entirety by two different laboratories (G. Shaw and L. Hood) with 100% concordance.
- **Epitope:** The amino acid sequence of the epitope of interest as defined in the reference, based on the reference strain used in the study defining the epitope. On rare occasions, when only the epitope location and not the actual epitope sequence was specified in the original publication, and the sequences were numbered inaccurately by the primary authors, we may have misrepresented the epitope's amino acid sequence. Therefore epitopes that were not explicitly written out in the text in the primary publication, those that we determined by looking up the reference strain and the numbered location, are followed by a question mark in the table.
- **Antigen:** The antigenic stimulus of the CTL response.
- **Species(HLA):** The species responding and HLA of MHC specificity of the epitope.
- **Reference:** The primary reference.

Following the entry for a given CTL epitope are brief comments explaining the context in which the epitope was defined. If the same epitope was studied in several labs, each study is cited in its own bulleted entry.

HIV CTL EPITOPES SORTED BY HLA RESTRICTING ELEMENT

This section is a table of the epitopes included in Section 1 that have known HLA restricting elements, organized by the restricting element. Anchor and auxiliary residues for HLA molecules are listed, and if anchor residues with appropriate spacing are evident in the epitope, they are emboldened and underlined. This table provides minimal information about the epitopes; for more information see the tables where epitopes are organized by protein location.

HIV PROTEIN EPITOPE MAPS:

Because of the increasing number of defined epitopes, only human CTL epitopes defined to within a region of 21 amino acids or less, with a known HLA specificity, are indicated on the HIV protein epitope maps.

The location and HLA restriction elements of CTL epitopes are indicated on protein sequences of the WEAU clone 1.60. These maps are meant to provide the relative location of epitopes on a given protein, but the WEAU sequence may not actually carry the epitope of interest, as it may vary relative to the sequence for which the epitope was defined. Epitopes are numbered in bold on the maps; the map numbering corresponding to the numbering of the epitope sequence alignments.

ALIGNMENTS:

As with the MAPS, only human CTL epitopes defined within a region of 21 amino acids or less, with a known HLA specificity, have corresponding alignments. For each numbered epitope in the epitope-protein maps, an alignment was generated from the protein sequence alignments in the HIV-1 genetic sequence database. All epitopes are aligned to the subtype B consensus (the most common amino acid found in subtype B in each position), with the sequence used to define the epitope indicated directly above the B consensus. In consensus sequences an upper case letter indicates the amino acid was present in all sequences, a lower case letter indicates the amino acid was present in most sequences in a given position, and a question mark indicates two or more amino acids were represented with equal frequency. In the master alignment files from which the epitope alignments were created, there are many partial sequences that do not span entire genes. Many of these partial sequences were removed before extracting the epitope alignments, but after the consensus sequences for the master alignment were calculated. As a result, the consensuses as shown in the epitope alignments correctly record the consensus of the master alignment of the database but not necessarily the consensus of the sequences shown in the epitope alignment. We used the full length proteins in the 1997 HIV-1 database protein alignments for this section. The alignments were modified to optimize the alignment relative to the defined epitope and minimize insertions and deletions. A dash indicates identity to the consensus sequence, and a period indicates an insertion made to maintain the alignment. Stop codons are indicated with a $, and frameshifts by a #; some epitopes are only partially sequenced.

REFERENCES AND NOTES